UNDERSTANDING THE GENETIC HISTORY OF THE MALAY AND ORANG ASLI POPULATIONS IN PENINSULAR MALAYSIA BY KILLER CELL IMMUNOGLOBULIN-LIKE RECEPTORS (KIR)

NUR WALIYUDDIN HANIS BIN ZAINAL ABIDIN

UNIVERSITI SAINS MALAYSIA

2018

UNDERSTANDING THE GENETIC HISTORY OF THE MALAY AND ORANG ASLI POPULATIONS IN PENINSULAR MALAYSIA BY KILLER CELL IMMUNOGLOBULIN-LIKE RECEPTORS (KIR)

by

NUR WALIYUDDIN HANIS BIN ZAINAL ABIDIN

Thesis submitted in fulfillment of the requirements

for the degree of

Doctor of Philosophy

March 2018

ACKNOWLEDGEMENTS

Alhamdulillah, all praise to Allah for giving me the chance to live and allow me to finish up my research and thesis. Thank you Allah for conferring me with understanding and supported families especially my mother, Rohani binti Mohd Ardzi, who never stop praying for my success. Not to forget, my beloved wife, Siti Nazariah binti Wahab, who willing to share all my busyness and hear all my problems throughout this six years of our PhD life. This PhD is dedicated for you. Thanks again for giving and raising three wonderful children of us; Nur Irdina Insyirah, Ahmad Farid Hilmi and Zahra Humaira.

This research will never been done without the guidance of all my supervisors especially my respected main supervisor, Assoc. Prof. Dr. Zafarina Zainuddin, who taught me not only to be a good researcher, but also how to be a good person in life. I was inspired by the way she taught me in class since my degree study in USM Kubang Kerian. She never neglected her students even being apart from us. There are no other words that can describe my deepest gratitude for all your guidance and thanks again for accepting me to be your student.

My great acknowledgement goes to all my three co-supervisors. Mr. S. Panneerchelvam, who always be with me in the lab, sharing all his knowledge about DNA and the basic concept of Science. He just likes my father in PPSK. Prof. Norazmi Mohd Nor, who always triggering me to think deeper in everything, correcting my English vocabulary, polish my presentation style and challenge me to become more than a lab technician. He is a real motivator of me. Not to forget, Dr

Edinur Hisham Atan, who responsible for my first publication. I will never forget all your guidance and assistance.

A special thanks to all the staffs in HID/DNA lab and Forensic lab, postgraduate fellows and all lecturers/staffs in PPSK to for their help and support. They are the unsung hero of me. Lastly, thanks to the USM for granting me a fellowship and Ministry of Higher Education for funding this research until the end.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

A	Adenine
AE	Elution buffer
AN	Austronesian
BAM	binary alignment map
bp	Base pair
BL	Lysis buffer
BSA	Bovine serum albumin
BW	Washing buffer
BWA-MEM	Burrow-Wheeler Aligner-Memory Efficient Mapping
C	Cytosine
ºC	Degree celsius
D	Domain
ddNTP	Dideoxynucleotide triphosphate
dH_2O	Deionized water
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide triphosphate
ECD	Enrichment Control DNA
EDTA	Ethylene Diamine Tetra-acetic Acid
e.g.	for example
ES	Enrichment system
EtOH	Ethanol
F	Gene frequency

FASTA	text-based DNA sequence format
FASTQ	text-based DNA sequence format with quality score
G	Guanine
GATK	Genome Analysis Toolkit
gDNA	Genomic DNA
gF	Estimated gene frequency
HLA	Human leukocyte antigens
HW	Hardy-Weinberg
ID	KIR genotype number
i.e.	this is/namely
Ig	Immunoglobulin
IGV	Integrative Genomics Viewer
IPC	Internal positive control
IPD	Immuno Polymorphism Database
ISP	Ion Sphere™ Particle
Kb	Kilo base
KIR	Killer cell immunoglobulin-like receptors
L	Long/Liter
LD	Linkage disequilibrium
LRC	Leukocyte receptor complex
M	Molar
MAF	Minimum allele frequency
$MgCl_2$	Magnesium chlorite
mL	Milliliter
mM	MilliMolar

MP	Malayo-Polynesian
mtDNA	Mitochondrial DNA
MΩ	Mega Ohm
nAA	Common Homozygotes
nAB	Heterozygotes
NaOH	Sodium hydroxide
nBB	Rare Homozygotes
ng	nanogram
NGS	Next-generation sequencing
$(NH_4)_2SO_4$	Ammonium sulfate
NK	Natural killer
NS	Nucleotide substitution
OA	Orang Asli
OT2	One Touch™ 2
P	Pseudo
PCA	Principle component analysis
PCR-SSP	Polymerase chain reaction-sequence specific primer
PGM	Personal Genome Machine
*p*mol/L	picomole/liter
QC	Quality control
rATP	Ribonucleoside-adenosine triphosphate
RE	Restriction enzyme
RNA	Ribonucleic acid
S	Small
SEA	Southeast Asia

SNP	Single nucleotide polymorphism
SOLiD	Sequencing by Oligo Ligation Detection
SSP	Sequence-specific primer
T	Thymine
TBE	Tris Borate EDTA
Tris-HCL	Tris hydrochloric acid
TW	Washing buffer
V	Volt
vcf	Variant Call Format
UCSC	University of California Santa Cruz
UPW	Ultrapure water
w/v	Weight/volume
x *g*	Times gravity
χ^2	Chi-square
µL	Microliter
µM	MicroMolar

MEMAHAMI SEJARAH GENETIK RUMPUN MELAYU DAN POPULASI ORANG ASLI DI SEMENANJUNG MALAYSIA MELALUI PENGENOTIPAN GEN *KILLER CELL IMMUNOGLOBULIN-LIKE RECEPTOR* (KIR)

ABSTRAK

Gen *killer cell immunoglobulin-like receptor* (KIR) berpotensi untuk dijadikan sebagai penanda molekul antropologi berdasarkan kepelbagaian kandungan gen ini dalam populasi manusia. Variasi yang terdapat pada jujukan alel gen ini juga mampu diaplikasikan sebagai pengenalan diri dengan memperlihatkan perbezaan antara dua individu yang dipilih secara rawak. Kajian sulung ini telah mengenalpasti 16 kandungan gen dan variasi jujukan gen *KIR* pada populasi Melayu dan Orang Asli (OA) di Semenanjung Malaysia, menggunakan kombinasi kaedah *polymerase chain reaction-sequence specific primers* (PCR-SSP), *next-generation sequencing* (NGS) dan disahihkan dengan penjujukan DNA Sanger. Sejumlah 333 individu mewakili 10 subetnik Melayu (Aceh, Banjar, Bugis, Jawa, Campa, Kedah, Kelantan, Mandailing, Minangkabau dan Patani) dan 167 individu mewakili enam suku OA (Batek, Kensiu, Lanoh, Che Wong, Semai dan Orang Kanaq) telah dipilih sebagai sampel kajian. Hasil kajian menemui 40 genotip *KIR* pada populasi Melayu manakala 25 genotip pada populasi OA. Secara umumnya, populasi Melayu memiliki kandungan gen *KIR* yang bersifat homogen dengan kecenderungan terhadap haplotip kumpulan *A* manakala OA lebih bersifat heterogen dengan mempamerkan kepelbagaian genotip dan kumpulan haplotip gen *KIR*. Analisis *principal component analysis* (PCA) memperlihatkan subetnik Melayu diplotkan di antara populasi Cina, India dan OA.

Suku Semang pula menunjukkan kesamaan genetik dengan populasi Afrika dan India serta populasi pribumi Papuan dan Australia. Malahan, semua populasi ini berkongsi genotip unik yang sama (*AB6*, *BB71*, *BB73* dan *BB159*) serta mempamerkan haplotip kumpulan *B* yang tinggi, dan seterusnya boleh dispekulasikan sebagai ciri-ciri umum manusia tamadun awal. Kaum Senoi menunjukkan haplotip kumpulan *A* yang tinggi serta mengekspresikan fenotip hibrid antara puak Indo-Cina dan kaum Semang. Sebaliknya, Orang Kanaq menampilkan perbezaan genetik dengan populasi lain termasuklah populasi Taiwan/Cina dan penutur bahasa Melayu-Polinesia. Aplikasi NGS telah berjaya mengenalpasti 302 alel daripada tujuh gen *KIR* populasi Melayu dan OA. Gen *KIR2DL1*, *KIR2DL4*, and *KIR3DL3* menunjukkan variasi alel yang rendah dan terhad kepada alel-alel tertentu. Hal ini boleh dikaitkan dengan persaingan semulajadi dan seterusnya cenderung untuk mengekalkan alel yang lebih kuat. Selain itu, gen *KIR2DS4*, *KIR3DL1*, *KIR3DL2* and *KIR2DL3* berevolusi dalam persaingan seimbang dengan memperlihatkan kadar polimorfisme alel yang tinggi. Gen *KIR3DL2* dilihat sebagai penanda terbaik untuk penyelidikan populasi genetik dengan mempamerkan taburan alel-alel unik (*KIR3DL2*001* dan **002*) yang menyokong teori migrasi manusia moden. Hasil penyelidikan ini telah berjaya menjana kerangka data *KIR* yang komprehensif serta mengaplikasikan kepelbagaian kandungan gen dan jujukan alel gen ini dalam merungkai sejarah genetik populasi dunia terutamanya bagi rumpun Melayu dan populasi OA.

UNDERSTANDING THE GENETIC HISTORY OF THE MALAY AND ORANG ASLI POPULATIONS IN PENINSULAR MALAYSIA BY KILLER CELL IMMUNOGLOBULIN-LIKE RECEPTORS (KIR)

ABSTRACT

Extensive diversity of killer cell immunoglobulin-like receptor (KIR) genes among diverse populations has made these immune genes become a potential anthropological marker. The variations of *KIR* allotypes between two randomly selected individuals are also useful in individual identification. Here, we genotyped *KIR* genes content and allelic polymorphisms for the first time for Malay and Orang Asli (OA) populations in Peninsular Malaysia using polymerase chain reaction-sequence specific primers (PCR-SSP), next-generation sequencing (NGS) and finally optimized by Sanger typing method. A total of 333 individuals representing 10 Malay subethnic groups (Aceh, Banjar, Bugis, Jawa, Champa, Kedah, Kelantan, Mandailing, Minangkabau and Patani) and 167 individuals from six OA subgroups (Batek, Kensiu, Lanoh, Che Wong, Semai and Orang Kanaq) have participated in this research. Fourty *KIR* genotypes were observed among Malays and 25 genotypes in OA populations. Generally, the Malays having homogenous haplotype profile with slightly high frequencies of haplotype *A*, whereas OA are heterogeneous with various genotypes and wide range of haplotypes. Based on the principal component analysis, the Malay subethnic groups are located between the Chinese, Indian and Orang Asli populations. The Semang subgroups show genetic affinity toward the Australian Aborigines, Papuan, Indian and African populations. They shared several unique genotypes (*AB6*, *BB71*, *BB73* and *BB159*) and have high frequencies of

Haplotype *B,* which could be speculated as general features of ancient populations. The Senoi demonstrated high frequency of haplotype *A*, and expressing hybrid phenotypes between Indo-China tribes and Semang population. In contrast, Orang Kanaq, the only Proto-Malay subgroup studied is significantly different from both related Taiwanese/Chinese and neighboring Malayo-Polynesian speaking populations and show evidence of becoming a distinct population. The NGS has successfully characterized 302 different allotypes of seven studied *KIR* genes among Malay and OA populations. *KIR2DL1*, *KIR2DL4*, and *KIR3DL3* genes show less allelic variations and in favor on particular allele, which possibly being affected by natural selection and creates selective sweep. In addition, *KIR2DS4*, *KIR3DL1*, *KIR3DL2* and *KIR2DL3* are highly polymorphic genes with considerable number of alleles discovered, suggesting that these genes evolved on balanced selection. The *KIR3DL2* gene could be the best marker for population studies, as it comprises signature alleles (*KIR3DL2*001* and **002*) that support the theory of modern human migration. Comprehensive datasets generated from this research demonstrate the value of *KIR* genes content and allelic polymorphisms in elucidating genetic relationships between the Malays and OA with other world populations.

CHAPTER 1

INTRODUCTION

1.1 Introduction

This research is to fill the gap of genetic history in Peninsular Malaysia by characterizing killer cell immunoglobulin-like receptors (KIR) genes content and allelic polymorphism among the Malay and Orang Asli populations. These genes have significant role in human health including transplantation compatibility, pathogenesis of infectious and susceptibility to autoimmune diseases (Rajagopalan and Long, 2005; Carrington and Martin, 2006; Parham, Norman, Abi-Rached, Hilton, *et al.*, 2012; Moretta *et al.*, 2014; Hollenbach and Oksenberg, 2015). Presently, 604 different *KIR* genotypes and 907 allelic variations have been reported (Robinson *et al.*, 2013; González-Galarza *et al.*, 2015) among global populations. Extensive diversity of *KIR* genes among different populations and even between two randomly selected individuals has made these functional genes to become potential genetic marker for population genetics (Rajalingam *et al.*, 2001; Shilling *et al.*, 2002).

1.2 Populations in Peninsular Malaysia

Malaysia is located at the heart of Southeast Asia (SEA) connecting the Asia mainland and islands of SEA. This country is populated by 28.7 million of multifarious society with different ethnics, racial, cultural and religious. The Malays constitute the majority (50.1%), followed by Chinese (22.6%), Indians (6.7%) as well as numerous aboriginal people (11.7%), including in Sabah and Sarawak (Department of Statistics Malaysia, 2017).

1.2.1 Orang Asli

The aboriginal people of Peninsular Malaysia are locally called as *Orang Asli* (OA). Currently, there are more than 178,000 OA individuals living in Peninsular Malaysia which represent ~0.5% of Malaysian population (Ministry of Rural and Regional Development). They are categorized into three major groups (Semang, Senoi and Proto-Malays) based on their economic activities, physical appearance and language (Hood, 2006).

1.2.1(a) Semang

Semang are the earliest inhabitants of Peninsular Malaysia since their arrival around 45,000-60,000 years ago (Barker *et al.*, 2002; Macaulay *et al.*, 2005; Soares *et al.*, 2008; Omar and Atoma, 2009). The Semang population only numbered around 5,000 individuals who belong to six subgroups; Batek, Jahai, Kensiu, Kintaq, Mendriq and Lanoh. Currently, they live in small tribes within the low-land rainforests at the northern part of Peninsular Malaysia and still practicing hunter-gatherer activities (Ministry of Rural and Regional Development; http://www.rurallink.gov.my). They are mutually related to the negrito peoples in the Philippines, Australian Aborigines and indigenous tribes of Papua New Guinea by expressing same phenotypes such as broad nose, short-stature, low cheekbone, frizzy hair and dark skin colour (Carey, 1976).

1.2.1(b) Senoi

Senoi is the largest group of OA in Peninsular Malaysia (~98,000 individuals) and consist of six subgroups; Che Wong, Jahut, Mahmeri, Semaq Beri, Semai and Temiar. The Senoi could be traced back from the Indochina during the Neolithic era

(7,000 years ago) based on mitochondrial DNA (mtDNA) analysis, language affinity (Austro-Asiatic language) and express similar phenotypes with the mountain tribes of Cambodia and Vietnam (Carey, 1976; Bellwood, 1993; Hill *et al.*, 2006). Their settlements are scattered across the high lands of central Peninsular Malaysia and until recently, still practicing slash and burn farming.

1.2.1(c) Proto-Malays

Proto-Malays is one of the Austronesian (AN) speakers. There are about 75,000 individuals who belong to six subgroups; Orang Kanaq, Orang Kuala, Orang Seletar, Jakun, Semelai and Temuan, currently living in Peninsular Malaysia (Ministry of Rural and Regional Development; http://www.rurallink.gov.my). Their physical appearance, language (Malayo-Polynesian) and lifestyle are closely similar to modern Malays (Deutro-Malays)(Kasimin, 1991). Most of the Proto-Malays are knowledgeable in oceanography and their settlements are concentrated along the coastal areas of southern Peninsular Malaysia. The Proto-Malays were previously seafarers, but are now mostly fishermen and merchants (Rohani and Nur Hidayah, 2014). However, the origins of the Proto-Malays are still obscure and open for debate due to contradictory data among language, archaeology and genetic studies.

1.2.2 Malays

Malays are Malayo-Polynesian speaking population subset of AN speaking population. Various Malay subethnic groups from wide range of SEA have migrated to Peninsular Malaysia as early as 14th century or even earlier due to trading activities, invasion and suppression by the European colonials, civil wars, infertile land and high population density in their homeland (Musa, 1992; Kato, 1997; Abdul

Fattah, Rusdi and Md.Akbal, 2011; Rahman, 2014). These episodes involved the people primarily from Sumatra (Aceh, Mandailing and Minang Malays), southern Kalimantan (Banjar Malays), Sulawesi (Bugis Malays), Cambodia (Champa Malays), Java Island (Jawa Malays) and southern Thailand (Patani Malays). Mix-marriage between the Malays and OA populations could be the basis of emerging of the Deutro-Malays. Admixture between Arabian, Chinese, European, Indian and Persian traders and missionary has further created complexity of genetic structure of the Deutro-Malays (Hussin, 2008).

1.3 Population genetics

Population is defined as a group of individuals living in the specific region and practices their own culture and language. Individuals within population may carry different alleles of many genes and even unique among distinct populations. The phenotypic variations such as skin colour, facial features, body size or hair type which are strikingly different among ethnic and racial reflected by their genetic structure. The diversity of genetic structure among different populations can be caused by genetic mutations, migration and environmental changes. Population genetic studies could build up a picture of human evolutions, trace the routes of migration and analyse the relationship between populations by utilizing the gene and allele of genetic markers. Among famous genetic markers used for population genetics are mtDNA, Y-chromosome and single nucleotide polymorphism (SNP) (Jinam, Phipps and Saitou, 2013; Soares *et al.*, 2016). Recently, the highly polymorphic immunogenetic markers such as human leukocyte antigens (HLA) and killer cell immunoglobulin-like receptors (KIR) have been growingly used for population study (Kulkarni *et al.*, 2008; Parham, 2008). These biomarkers are not

only for analysing disease susceptibility, but also potentially become a new signature for population genetics.

1.4 Killer cell immunoglobulin-like receptors

KIR are one of the various receptors expressed on the surface of natural killer (NK) cells and subsets of T lymphocytes (Lodoen and Lanier, 2006). The *KIR* provide immune defence through recognition with major histocompatibility complex (MHC) class I molecules expressed on the cell surface of most human cells (Parham, Norman, Abi-Rached and Guethlein, 2012). The *KIR* are encoded by a family of highly polymorphic and homologous genes clustered tandemly arrayed within 150 Kb of leukocyte receptor complex (LRC) region on chromosome 19q13.4 (Vivier and Daëron, 1997) (Figure 1.1). Nomenclature of *KIR* genes is designed based on the number of extracellular immunoglobulin (Ig) domains (D) and the length of cytoplasmic tail; short (S) or long (L). The *KIR* genes with S cytoplasmic tail express immunoreceptor tyrosine-based activating motifs to initiate NK cells activity by interaction with adaptor molecules. In contrast, L cytoplasmic tail conveys immunoreceptor tyrosine-based inhibitory motifs to induce inhibitory reaction (Selvakumar, Steffens and Dupont, 1996; Lanier *et al.*, 1998). Numerical designation for *KIR* allele assignment is similar fashion to the *HLA* allele nomenclature (Marsh *et al.*, 2003). The first three digits after an asterisk (*) indicate variants in the exon parts which lead to non-synonymous changes. The next two digits indicate variants in the exon parts leading to synonymous changes. The last two digits correspond to the variants in an intron, promoter or other non-coding regions (Figure 1.2).

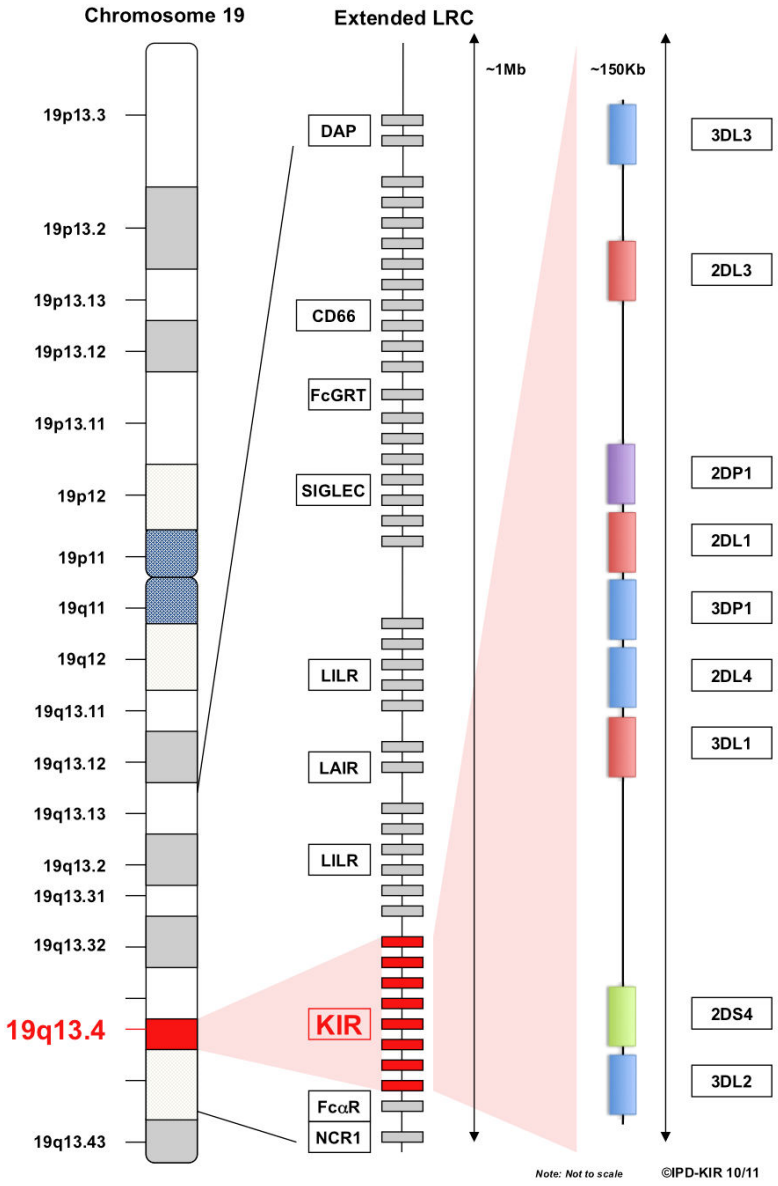

Figure 1.1: *KIR* genes loci encoded within 150 Kb stretch of the 1 Mb long extended Leukocyte Receptor Complex on chromosome 19q13.4 (adapted from http://www.ebi.ac.uk).

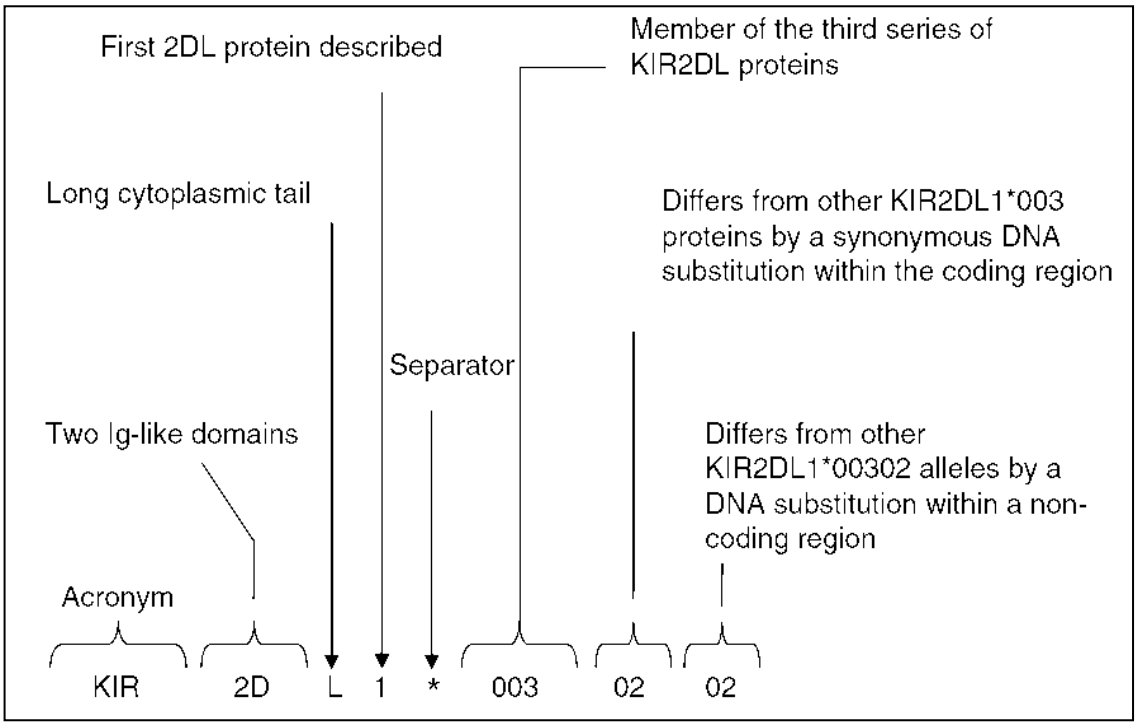

Figure 1.2: Designation of *KIR* genes and alleles nomenclature (adapted from Middleton and Gonzelez, 2010).

1.4.1 *KIR* genes

There are 15 *KIR* genes loci including two pseudogenes (*KIR2DP1* and *KIR3DP1*) have been identified so far with consideration of *KIR2DL2*/*KIR2DL3* and *KIR3DL1*/*KIR3DS1* as allele of same locus (Uhrberg *et al.*, 1997; Norman *et al.*, 2007; Parham, Norman, Abi-Rached and Guethlein, 2012). Six genes are involved in activating reaction namely *KIR2DS1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS4*, *KIR2DS5* and *KIR3DS1*. The other eight genes are participating in inhibitory reaction namely as *KIR2DL1*, *KIR2DL2*, *KIR2DL3*, *KIR2L5A*, *KIR2DL5B*, *KIR3DL1*, *KIR3DL2* and *KIR3DL3*. There are also possibilities of one gene (*KIR2DL4*) that can transduce both types of signals.

1.4.2 *KIR* haplotypes and genotypes

KIR genotypes are designated based on the combination of *KIR* genes present in an individual. They are organised into *A* and *B* haplotypes based on the number and type of *KIR* genes (Uhrberg *et al.*, 1997). Both haplotypes consist of four fixed genes called framework genes:- *KIR2DL4*, *KIR3DL2*, *KIR3DL3* and *KIR3DP1*. Haplotype *B* is recognized by the presence of one or more of the following genes; *KIR2DL2*, *KIR2DL5*, *KIR2DS1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS5* and *KIR3DS1*. Haplotype *A* is less variable and is classified by the presence of fixed inhibitory receptors (*KIR2DL1*, *KIR2DL3* and *KIR3DL1*) along with the activating (*KIR2DS4* and *KIR2DL4*) receptors (Uhrberg *et al.*, 1997). To date, 604 unique *KIR* genotypes have been discovered among 160 populations in Allele Frequency Net Database (González-Galarza *et al.*, 2015).

1.4.3 *KIR* allelic polymorphisms

In addition to *KIR* genes content, allelic polymorphisms appear between individual by exhibiting high degree of nucleotide sequence variations which affects the levels of cell surface expression, binding affinities and NK cell effector function (Middleton and Gonzelez, 2010). As for July 2017, 907 *KIR* alleles have been characterized and deposited in Immuno Polymorphism Database (IPD) (Robinson *et al.*, 2013). The allelic variation tends to occur all over the gene due to unequal crossing over between alleles of a single locus and separate gene loci. It is also suggested that much allelic variation is derived from non-reciprocal recombination between two *KIR* genes and also influenced by the sequence polymorphisms of *HLA* genes (Shilling *et al.*, 2002; Yawata *et al.*, 2006).

1.4.4 *KIR* in population genetics

Recently, there has been a substantial increase in the number of *KIR* genotypes and alleles reported across geographically and ethnically defined populations. East Asian regions such as China (Yao *et al.*, 2011; Shi *et al.*, 2013) and Japan (Yawata *et al.*, 2006) exhibit high frequencies of haplogroup *A* while African (Williams and Middleton, 2004; Middleton *et al.*, 2008), Indian (Rajalingam *et al.*, 2002, 2008), Australian Aborigines (Toneva *et al.*, 2001) and Papuan (Velickovic, Velickovic and Panigoro, 2009) populations express high frequencies of haplogroup *B*. It has also been speculated that haplotype *B* may be generally predominant among ancient population (Middleton and Gonzelez, 2010; NurWaliyuddin *et al.*, 2015).

1.5 Significance of study

The use of genetic markers for delineating the phylogeography and ethnogenes are more focused on mtDNA, Y-choromosome and SNP analyses. Here, we manipulate this functional gene (*KIR* genes) as an alternative marker for population genetics. The characterization of *KIR* gene and allele polymorphisms among OA and Malay populations will provide a wider perspective of population genetics in SEA region. With extensive variability of *KIR* system across ethnically defined populations, it is interesting to determine the distribution of *KIR* genotypes and haplotypes in the OA subgroups and Malay subethnic groups in order to help delineate their genetic relationship and migration history in this region. In addition, *KIR* exome sequencing by NGS could further increase the discriminatory power of *KIR* systems for population genetic study by providing accurate measurement of genetic variation.

1.6 Problem statements

The history and origin of Malay and OA populations are still being debated among various fields of studies including linguistic, anthropology, archaeology, and genetics. This study could provide supporting information to help in tracing the origin of these populations by manipulating the *KIR* datasets of Malay and OA populations, while providing baseline information for future health related issues. Hitherto, there is no available data for *KIR* genes content and alleles for Malaysian populations. This is the first ever study for *KIR* for OA and Malay populations.

The studied OA populations are believed to suffer severe genetic drift and they are exposed to several threats that may lead to their extinction. Findings from the present study are expected to become increasingly important due to the increasing number of

studies showing associations between specific *KIR* genotypes and various diseases such as diabetes mellitus (Santin *et al.*, 2006), hepatitis B (Lu *et al.*, 2008) and malaria (Taniguchi and Kawabata, 2009), which are commonly found among OA (Sumithran and Prathap, 1977; Ali, Shamsuddin and Khalid, 1991; Jamaiah *et al.*, 2006). This study is necessary to understand the role of *KIR* genes, if any, for their survival.

1.7 Objectives of study

1.7.1 General objective

To characterize 16 *KIR* genes content and determine the allelic variations within exons of seven most polymorphic *KIR* genes among OA subgroups and Malay subethnic groups in Peninsular Malaysia using PCR-SSP technique and NGS technique.

1.7.2 Specific objectives

i) To report *KIR* genotype and haplotype of Malay and OA populations as no previous molecular population genetic studies involving these populations.

ii) To delineate genetic relationship and migration history of Malay and OA populations with the global relative populations.

iii) To provide strong foundation for further investigations of the prehistoric dispersal, evolutionary pattern while providing baseline information for future health related issues by using *KIR* allelic variation.

iv) To optimize NGS bioinformatics pipelines for low sequence coverage using Sanger typing method.

CHAPTER 2

LITERATURE REVIEW

2.1 History of Southeast Asia

Southeast Asia (SEA) is a home for 620 million of AN speaking populations consisting 11 countries; Cambodia, Laos, Myanmar, Thailand, Vietnam and Peninsular Malaysia in the mainland while Brunei, East Malaysia, East Timor, Indonesia, the Philippines and Singapore in the island of SEA. The SEA is the passage for human migrations through the region and into the Pacific region and Australia continent as well as connecting between the Chinese and Indian civilizations. Before the last glacial maximum (~19,000 years ago), southern shelf of Thailand, Peninsular Malaysia, islands of Sumatra, Java and Borneo were joined to the Asia mainland as a vast continuous lowland tropical rainforest known as Sundaland (Figure 2.1). Several climatic changes took place around 14.5k, 11.5k and 7.5k years ago had gradually rise the sea-level by ~120 meter from the present day and submerged the vast lowland of ancient Sundaland (Pelejero *et al.*, 1999; Bird, Taylor and Hunt, 2005). Thus, thousands of islands were created as today and brought a great impact for the mankind of this region. These demographic changes slowly stimulated a highly maritime-oriented culture between the SEA populations which later promoting linguistic, cultural diffusion and assimilation. Rivers and sea became the important routes connecting the surrounding regions for trading, searching for new settlement and religious activities. Eventually, many Malay maritime kingdoms were established near the rivers to control the sea trade activities.

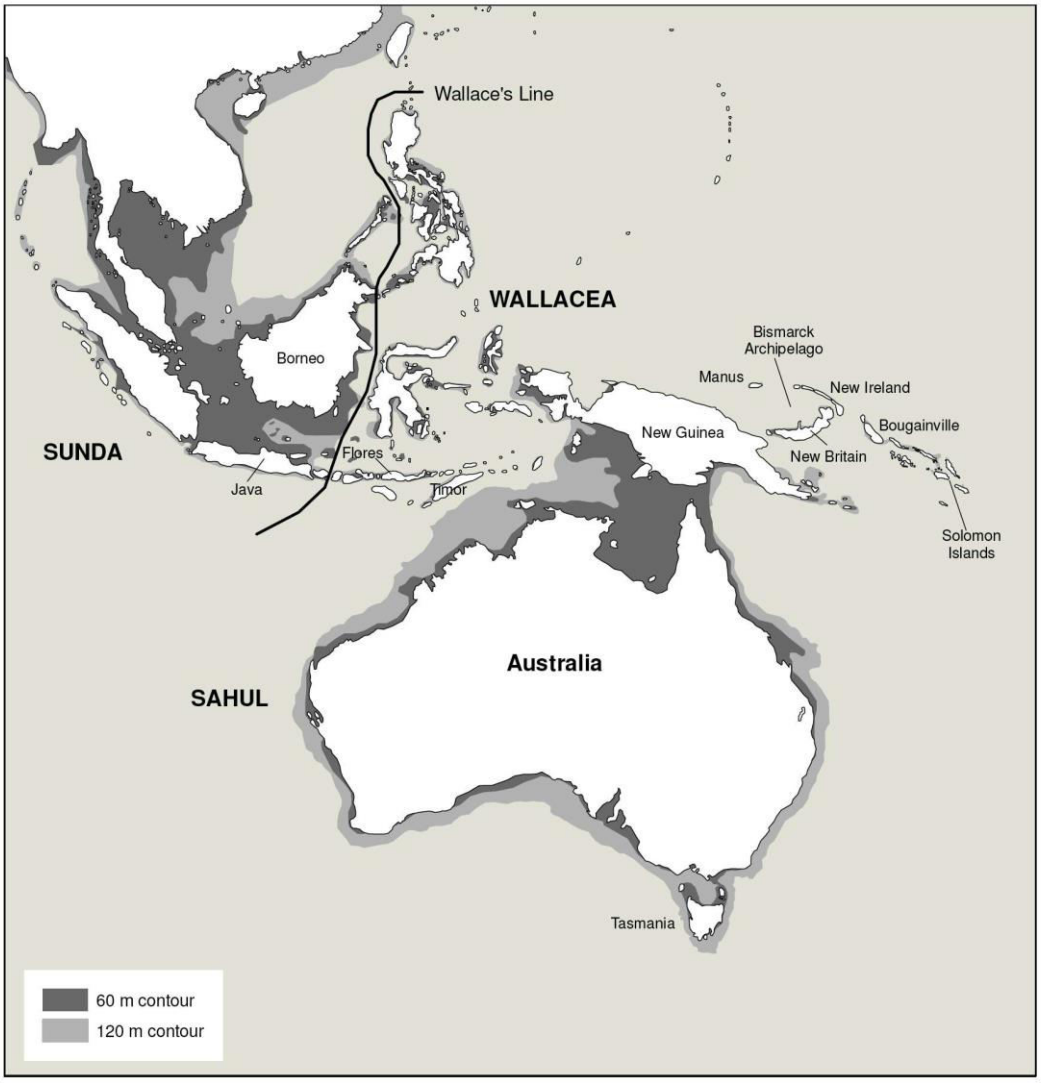

Figure 2.1: The map of ancient Sundaland (grey contour) which comprise of southern shelf of Thailand, Peninsular Malaysia, islands of Sumatra, Java and Borneo before the last glacial maximum. White contour shows the current map of SEA and Australia continent (adapted from https://sergejohn.wordpress.com).

2.2 History of the Malays

Malay civilization was possibly begun before century with the emergence of Funan and Champa kingdoms in Indochina, Langkasuka in southernmost province of Thailand as well as Kedah Tua kingdom in Peninsular Malaysia (Mahmud, 1999; Hussin, 2004). In 2nd century, Ptolemy, Egyptian astronomer has portrayed the word *Melayu* in his Golden Khersonese as 'Melayu Kulon'. *Melayu* was also written in Chinese records during Fa Hsien's expedition in 5th century which referring to Palembang, Sumatra. Furthermore, Marco Polo has identified 'Malayur' as ancient kingdom in Sumatra during his expedition around 1290 (Wahyudin, 2014). These are all among the evidences of existence of Malay kingdoms.

Linguistically, Malay language is categorized under western Malayo-Polynesian (MP) language, a subgroup of AN language family. Beside Malay language, Malagasy, Indonesian, Javanese, Sundanese, Madurese, Balinese and Tagalog languages are the also categorized under western MP languages, which are spoken by approximately 200 million people of SEA (Green, 1966; Blust, 2013).

The Malays are widely known since the establishment of *Kesultanan Melayu Melaka* (Malacca Sultanate) where the Malay language became the lingua franca, a commercial language spoken by traders all over the world (Omar and Atoma, 2009; Din, 2011). Many traders and preachers from Arab, China, Europe, India and Iran had built their settlements in Peninsular Malaysia while waiting for the monsoon change. Greater or lesser admixture between the traders and local communities has diversified the genetic structure of Malay population (Hussin, 2008). There was also influx of neighbouring Malays occurred as early as 14th century due to invasion and

suppression by the European colonials, civil wars, infertile land and high population density. These episodes created various Malay subethnic groups in Peninsular Malaysia mainly from Sumatra (Aceh, Mandailing and Minang people), southern Kalimantan (Banjarese), Sulawesi (Buginese), Cambodia (Cham), Java Island (Javanese) and southern Thailand (Patani) (Figure 2.2).

2.2.1 Acehnese

Acehnese are people originating from the northernmost of Sumatra which ruled by Kingdom of Aceh Darussalam (1496-1903). Aceh kingdom achieved its golden era in 17th century expanded outward by sea and become epicenter of Islamic traders after the fallen of Malacca Sultanate (Din, 2011). Acehnese traders and religious scholars frequently visited Peninsular Malaysia and some continue to live with the local community in northern Malaysia especially in Yan, Kedah and Langkawi island for at least several decades.

2.2.2 Banjarese

Banjarese are Muslim Dayak living in coastal region of south Kalimantan controlled by Banjar Kingdom (1526 -1905) (Noor, 2012). They are well-known pepper producer and trader in south coast of Kalimantan. Since 1809, the Dutch has launched serial attacks and forced these people to retreat and seek safer places in Jambi, Sumatra and Peninsular Malaysia (Wahyudin, 2014). In Peninsular Malaysia, they are more concentrated in Kuala Selangor and Sabak Bernam (Selangor), Krian (Perak) and Batu Pahat (Johor). Until recently, Banjar Malays still preserve their culture and dialect by practicing marriage within Banjar people.

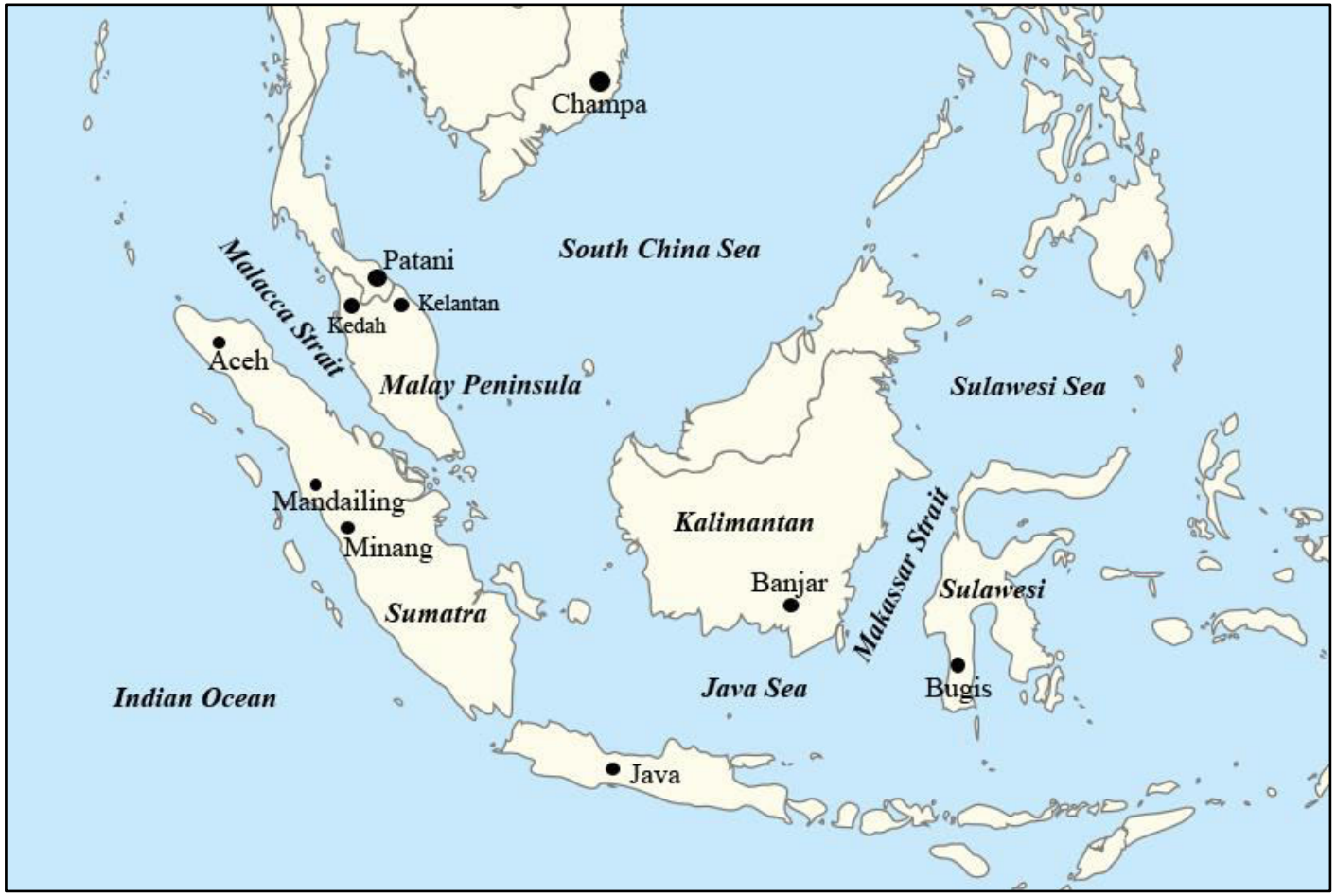

Figure 2.2: The map of Southeast Asia depicting the origins of various Malay subethnic groups.

2.2.3 Buginese

Bugis people were originated from the coastal area in southwestern of Sulawesi Island. Sailing and trading were the tradition of Buginese. Mass migration of Buginese from south Sulawesi to Borneo, Indragiri and Peninsula Malaysia started between 17th to early of 18th century due to civil war between Makasar and Bone Bugis. The involvement of the Dutch in this civil war had induced Perang Makasar in 1669 and forced Bugis people to move out (Abdul Fattah, Rusdi and Md.Akbal, 2011). In Peninsular Malaysia, most of the Bugis Malays settled in Johor, Selangor and Pulau Pinang.

2.2.4 Cham people

Champa is one of the earliest Malay kingdom governing the central and southern coastal part of Vietnam lasting from 2nd to 17th century (Setudeh-Nejad, 2002; Lombard, 2011). The name of South China Sea was once known as Champa Sea during the fame of Cham kingdom. The expansion of Cham kingdom has triggered a conflict with Dai Viet (Vietnam) and caused long series of civil war. Started in 15th century, Dai Viet has launched serial invades and conquered several Champa territories (Musa, 1992). Once again in 1975, communist regime under Pol Pot's reign started the persecution of Muslim Chams with more than a million people have left their home (Mupiza, 2016). It was estimated more than 10,000 of Cham refugees were allowed to resettle in Malaysia (Ken, 2013). Nonetheless, the historical Islamic relationship between Chams and Malays had started long time ago with the built of *Masjid Kampung Laut* in Tumpat by Chams during 18th century. Numerous Champa Malays settled in Pengkalan Chepa and Gong Chepa, Kelantan since several hundred years ago (Yaacob *et al.*, 2010).

2.2.5 Javanese

Javanese is the largest ethnic group in Indonesia concentrated in Java Island. They achieved their glory during the peak of Majapahit Empire (1294-1527) (Mas'ud, 2008). The rapid growth of Java population has forced them to migrate out. However, there is no proper evidence of Javanese settlement in Peninsula Malaysia before 19th century. Only in 19th century, the British had recruited Javanese immigrants into Peninsular Malaysia as contract agriculture workers (Hugo, 1993). In addition, outnumber of Chinese and Indian immigrants have raised the concern among Malay nationalists. During the first decade of Independence, Malaysia government welcomed the Indonesian immigrants to strengthen the position of the Malay in Malaysia (Spaan, Van Naerssen and Kohl, 2002). In Peninsular Malaysia, most of Jawa Malays resided in Johor, Selangor and Perak state.

2.2.6 Kedah Malays

Kedah Tua (1st - 14th century) is one of the oldest Malay kingdoms in Peninsular Malaysia based on the archaeological evident of more than 80 royal heritage sites in Lembah Bujang (Samsudin *et al.*, 2010). Kedah Tua was also claimed as the most advanced ancient kingdom with the discovery of iron smelting technology aged from 3rd to 5th century (Akma, Saidin and Abdullah, 2011). In between 3rd to 13th century, Kedah Tua became an important trading port and provide an alternative inland route via Kedah-Yarang trans-peninsula connecting between Bay of Bengal and Gulf of Thailand (Samsudin *et al.*, 2010; Ramli and Rahman, 2012). The influence of Indian civilization was very obvious in Kedah with more than 50 artifacts of Hindu/Buddha temples built in between 4th to 12th century were discovered in Lembah Bujang (Arkib.gov.my, 2018).

2.2.7 Kelantanese

The strategic geographical location in the Gulf of Thailand and also the nearest land from the Cape Ca Mau (Vietnam) have made Kelantan as an appropriate meeting place for traders mainly from China, India as well as Middle East (Hussin, 2004). Cultural shifting between diverse populations has resulted Kelantan as a culturally rich state in Malaysia. Kelantan was also known as *Serambi Mekah*, an Islam propagation center with the built of many *sekolah pondok* (Islamic school) in this state. The spreading of Islam in Kelantan started before 12th century which is much earlier than any other Malaysia states (Mohd Nor *et al.*, 2012).

2.2.8 Mandailing people

Mandailing people are subethnic of Batak cluster originated from southwestern corner of North Sumatra. The history of Mandailing people is very scanty. The word 'Mandailing' was written in the Javanese epic of Nagarakretagama (1365) as one of the region controlled by Majapahit kingdom. Large migration of Mandailing people into Malaysia and Singapore occurred after the invasion of the Dutch in 1830 (Ansyah, 2012). In Malaysia, most of Mandailing settled in south of Perak and they contributed important role for the development of Perak kingdom (Lubis and Khoo, 2003).

2.2.9 Minang people

Minang is the largest ethnic group in West Sumatra governed by Minangkabau kingdom in Nagari Pagaruyung. This kingdom achieved its glory in 14th and 15th centuries and last until the capture of Sultan Pagaruyung by Dutch colonial in 1833 (Mahmud, 2014). They are well known for their matrilineal system called *adat*

perpatih, delineating descent and inheritance through the maternal line. Minang people migrated to Peninsular Malaysia mainly to Negeri Sembilan as early as 15th century during the fame of the Melaka kingdom. Succeeding waves of Minang migration occurred after the Dutch colonial attacked their homeland (Kato, 1997; Aslinda, Noviatri and Reniwati, 2015).

2.2.10 Patani people

Patani was part of Langkasuka kingdom (1st-13th century) located at southernmost province of Thailand including Satun, Songkhla, Yala, Narathiwat, Kelantan and Terengganu (Mahmud, 1999; Yaakob, 2012). Patani kingdom arose in 14th century as the continuation of Langkasuka. Culturally, Patani is one of the Malay principalities until the Siam and British sealed Anglo-Thai Treaty (1909) where British recognized Thailand's sovereignty over Patani (Aphornsuvan, 2003). Patani was once under controlled of Srivijaya and Majapahit kingdoms in eighth century (Yaakob, 2012) which gave rise to the entrance of Javanese people into Patani. The spreading of Hindu/Buddha in early century by Indian people has also contributed to the genetic makeup of Patani people. Thus, genetic composition of Patani people are blended with Indian, Javanese and Thailand people.

2.3 History of Orang Asli in Peninsular Malaysia

2.3.1 Semang

Archaeological evidences have traced the earliest human activities in Peninsular Malaysia at Kota Tampan, Lenggong in Perak state which are related to Paleolithic Age (Taha, 1989). These could be linked to the Semang population, the descendants of the earliest inhabitant of Peninsular Malaysia since their arrival ~45,000-60,000

years ago (Barker *et al.*, 2002; Omar and Atoma, 2009). They are also related to the first wave of modern humans migrating out of Africa travelling along southern coast of Arabian Peninsula towards India, SEA (Peninsular Malaysia) and continued moving southern towards Pacific region and Australia (Hill *et al.*, 2006; Soares *et al.*, 2008; Oppenheimer, 2012).

Currently, they are the smallest OA population with only numbered around 5,000 individuals compared to Senoi and Proto-Malays. The Semang are further divided into six subgroups based on their unique dialects (i.e. Batek, Jahai, Kensiu, Kintaq, Mendriq and Lanoh). Most of them are living in small tribes and practicing semi-nomadic lifestyle which depends on the seasonal bounties of the low-land rainforests at the northern part of Peninsular Malaysia. Their physical appearances look rather like the aborigine peoples in Indonesia, the Philippines, the Australian Aborigines as well as natives of West Irian Jaya and Papua New Guinea. The Semang are generally small in stature about five feet or less. They have varies skin colour from dark copper to black darker skin. Usually, they have low cheek-bones, broad noses, and typically wooly or frizzy hair which are similar to negrito phenotypes. Their languages are categorized under Austro-Asiatic (Aslian) language, a branch of Mon-Khmer language family, which are believed to be introduced via social interaction with the Senoi people.

2.3.1(a) Batek

The Batek, with a population of about 1,478 in 2010, is the second largest Semang subgroup with annual percentage growth rate of 4.77% for the period 1969-2010 (Table 2.1). The nomadic Batek inhabited mostly the primary and old secondary rain

Table 2.1: Populations size and annual population growth rate of six studied Orang Asli subgroups for the period 1969-2010.

Orang Asli Populations	Subgroups	Year			Annual population growth rate (%)		
		1969	2003	2010	1969-2003	2003-2010	1969-2010
Semang	Batek	500	1255	1,478	4.44	2.54	4.77
	Kensiu	100	232	221	3.88	-0.68	2.95
	Lanoh	260	350	360	1.02	0.41	0.94
Senoi	Semai	15,506	43,892	49,427	5.38	1.8	5.34
	Che Wong	272	664	579	4.24	-1.83	2.75
Proto-Malays	Orang Kanaq	50	83	139	1.94	9.64	4.34

Annual percentage growth rate = $(N_{present} - N_{past}) / N_{past} \times 100 \div Y$, where: N = population size and Y = number of years. Data are retrieved from (Carey, 1976; Nicholas, 2006; Omar *et al.*, 2014).

forest in southern Kelantan, northern Pahang, and western Terengganu (Masron, Masami and Ismail, 2013). They involved in a wide-range of economic activities including hunting, gathering, occasional wage labour, and bartering or selling forest products.

2.3.1(b) Lanoh

Another comparatively small tribe of Negrito is the Lanoh, showing low annual percentage growth rate of 0.94% over the past 41 years, numbering about 360 people in 2010 (Isa, 2014) (Table 2.1). Lanoh usually found in Lenggong and Gerik at upper Perak and they are suggested to be related to the discovery of Palaeolithic site of Kota Tampan (75,000-30,000 years ago) and the skeleton of peninsular Malaysia's oldest burial, Perak Man, which dated between 10,000 - 11,000 years ago (Saidin and Tacon, 2011; Isa, 2014). Their economic activities include rattan handicrafts, growing hill rice, lime, vegetables and tapping rubber tree (Isa, 2014). At present, individuals with pure Lanoh blood line are very few among their young generation due to intermarriage with Senoi subgroups (i.e. Temiar).

2.3.1(c) Kensiu

Kensiu is the smallest and considered as an endangered subgroup of Semang population with only 221 living individuals reported in 2010 (Nicholas, 2006; Omar *et al.*, 2014). They are having negative annual percentage growth rate (-0.68%) between 2003-2010 (Table 2.1). Majority of Kensiu reside in Lubuk Legong, Baling in Kedah and the rest could be found in northern Perak. Their socioeconomic status has increased with the developments of Ulu Legong Hot Springs. They are more socialize compared to Batek and Lanoh, and surrounded by the Malay communities.

However, assimilation with the Malays had resulted severe threat to their culture and language (Ghani and Lah, 2015).

2.3.2 Senoi

Senoi is the largest group of OA in Peninsular Malaysia (~98,000 individuals) and consist of six subgroups; Che Wong, Jahut, Mahmeri, Semaq Beri, Semai and Temiar. They migrated to Peninsular Malaysia during the Neolithic age ~7,000 years ago (Bellwood, 1993; Hill *et al.*, 2006). The Senoi is mutually related to Cambodia and Vietnam mountain tribes based on the language affinity (Diamond and Bellwood, 2003). They are also expressing similar physical features like Mongoloid people by having wavy hair, more pronounced cheekbones, and lighter skin color. Their settlements are scattered across the high lands of central Peninsular Malaysia (Bellwood, 1993). Nowadays, the Senoi involved in various wage sector and managing permanent agricultural activities (e.g. tea, cocoa farms, oil palm and rubber estates). However, there are few subgroups that are still practicing shifting cultivation and dependent on the forest for hunting, fishing and gathering wild fruits and jungle products (Nicholas, 2006).

2.3.2(a) Semai

Semai is the largest OA subgroup, of about 50,000 individuals which make up 30% of the OA population in Peninsular Malaysia. They are mainly found at the flank of the Banjaran Titiwangsa in northwest Pahang and south Perak (Omar *et al.*, 2014; Ghani, 2015). Traditionally, Semai supplementing their diet by growing hill rice, hunting and gathering. Nowadays, they are also planting cash crops such as rubber, oil palm or cocoa as well as trading forest produce (Dentan, 2004). Genetic study